AD-A071 969 NAVAL POSTGRADUATE SCHOOL MONTEREY CA F/G 5/9

CULTURE-FREE TESTING FOR SELECTION OF RECRUITS.(U)

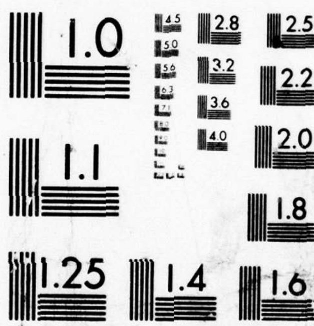
JUN 79 V A SHERMAN

UNCLASSIFIED NL

1 OF 1

AD
A071969

END
DATE
FILMED
8--79
DDC


1.0
4.5
5.0
5.6
6.3
7.1
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

NAVAL POSTGRADUATE SCHOOL

Monterey, California

THESIS

CULTURE-FREE TESTING FOR
SELECTION OF RECRUITS

by

Vining Alden Sherman, Jr.

June 1979

Thesis Advisor: James K. Arima

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)

6 Culture-Free Testing for Selection of Recruits.

5. TYPE OF REPORT & PERIOD COVERED

9 Master's Thesis, June 1979

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

10 Vining Alden/Sherman, Jr

8. CONTRACT OR GRANT NUMBER(s)

12 66p.

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Naval Postgraduate School
Monterey, California 93940

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS

Naval Postgraduate School
Monterey, California 93940

12. REPORT DATE

11 Jun[redacted]979

13. NUMBER OF PAGES

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

Naval Postgraduate School
Monterey, California 93940

15. SECURITY CLASS. (of this report)

Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Culture-free
Arima-Young Test
Testing
Recruit Testing

251450 Gu

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The purpose of this research was to compare a two-dimensional polygon discrimination learning (DL) test designed by Arima and Young with the Armed Services Vocational Aptitude Battery (ASVAB) currently being used for military selection.

A sample of 65 high school students who had taken the ASVAB in the military's High School Testing Program were given the DL test during the first two months of 1979. Little correlation

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
(Page 1) S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Gur

was found between the two tests. Multiple regression with IPR as the dependent variable indicated that the subtest General Information (GI) predicted Information Processing Rate (IPR) score to some extent. For whites only, the Armed Forces Qualification Test (AFQT) also served to predict IPR score. Results indicate that the DL test measures some other aspect of ability other than that demonstrated on the ASVAB and may therefore be of use in recruit screening.

Approved for public release; distribution unlimited

Culture-Free Testing for Selection of Recruits

by

Vining Alden Sherman, Jr.
Lieutenant, United States Navy
B.S., United States Naval Academy, 1972

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
June 1979

Author Vining A. Sherman, Jr.

Approved by: Thesis Advisor

Richard S. Elster

Second Reader

Chairman, Department of Administrative Sciences

Dean of Information and Policy Sciences

ABSTRACT

The purpose of this research was to compare a two-dimensional polygon discrimination learning (DL) test designed by Arima and Young with the Armed Services Vocational Aptitude Battery (ASVAB) currently being used for military selection.

A sample of 65 high school students who had taken the ASVAB in the military's High School Testing Program were given the DL test during the first two months of 1979. Little correlation was found between the two tests. Multiple regression with IPR as the dependent variable indicated that the subtest General Information (GI) predicted Information Processing Rate (IPR) score to some extent. For whites only, the Armed Forces Qualification Test (AFQT) also served to predict IPR score. Results indicate that the DL test measures some other aspect of ability other than that demonstrated on the ASVAB and may therefore be of use in recruit screening.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIG.

I. INTRODUCTION

The advent of the all-volunteer force in 1973 dictated many changes in personnel management procedures and conditions within the military. The largest, most immediate effect was that there no longer existed a massive pool from which to pick individuals meeting high, rather arbitrarily defined standards of mental and physical ability. In a very short time, the military services have had to shift from the luxury of administratively selecting from an unlimited source to the requirement to first create a source through more pronounced recruiting and then to select a slightly smaller number of entrants from this much smaller pool.

The basic personnel procurement problem facing the military is that it has been able only with great effort to meet the numerical requirements for new recruits. The present Chief of Naval Operations, Admiral Hayward, recently estimated the petty officer shortage in the Navy to be reaching 30,000. He also stated that this shortfall could possibly result in unfilled committments (Sinaiko, 1978). There are myriad reasons offered for manning shortfalls ranging from the passing of the baby boom to the increasing unattractiveness of military work (Cooper, 1978; Defense Manpower Commission, 1976).

There are numerous ways which present and projected shortfalls can be combatted. Retaining more present members of

the armed services is perhaps one of the best as well as one of the most difficult. Allowing more women into the service is another obvious means. Mental, physical, and moral guidelines could be relaxed and immediately alleviate recruiting problems. There is an ever-increasing list of alternatives for manning the armed forces, each with a benefit and each with a price.

There is one alternative, however, which is less generally considered. That is to improve our present selection procedures by devising more effective mental-testing programs. The purpose of the following research was to study such a possible improvement in the services' enlistment testing.

Traditionally the measure of an individual's mental ability has been taken through the use of various paper-and-pencil, verbally oriented tests. The present forms of the Armed Services Vocational Aptitude Battery (ASVAB) is one such paper-and-pencil test. This battery, when used alone to screen applicants for mental ability, suffers by its design limitations. Estes (1974) points out that an individual's successful passing of a verbal-type test provides useful information as to the extent to which the various prerequisites for successful performance have been simultaneously satisfied by the individual's combination of inherent capacities and past experiences. But failure gives little information because there is no indication as to why the individual failed. There are, unfortunately, many more reasons other than limited ability.

Individuals can possess high mental capability and still do poorly when tested in the traditional manner if they have not received the background and training necessary (and assumed by the examiner). Words used in the test may be strange; the subject matter may be foreign; the general format of the test may be new; an individual may give an answer which is meaningful in his own frame of reference but not within that of the examiner. Two people with equal mental potential can make impressively different scores on a test of mental ability if only one has been "taught" to take that type of test. Thus present tests would seem to be unfair to the bright but unorthodox person, to the culturally disadvantaged, and to the naive individual who lacks experience in taking standardized tests (Holtzman, 1971).

Since the ASVAB is one such standardized test, and since it is the sole means of identifying those "mentally qualified" for duty, it is quite conceivable that the armed forces are turning away a substantial group of otherwise qualified young men and women simply because of the format of the test. Since, also, the culturally disadvantaged and the inexperienced test takers tend to be nonwhite in this country, equal opportunity requirements are perhaps not being completely satisfied.

The Equal Employment Opportunity Commission (EEOC), established to enforce the Civil Rights Act of 1964, has looked fairly critically at employment testing in recent years. The important relevant point in the commission's guidelines is

that employment tests should be used only when there is demonstrated evidence of their validity for the purpose, and that tests must not discriminate unfairly against minority groups (Federal Register, Vol. 36, 1971). Much progress in both mental measurement and validation procedures must be realized before these guidelines can be satisfied. Minorities still score lower on service entrance examinations than whites although it has yet to be shown that nonwhites perform less well on the job (Lockman, 1976; Bilinski, 1974). It would be well for the services to resolve this incongruity of their own volition since there will probably be pressure from outside to do so in the future.

The research presented here was undertaken to study one alternative means of testing mental capabilities of potential recruits. The test was originally designed by Arima (1978) as a test of learning ability. It is a culture-fair or culture-free approach in that an attempt has been made to limit penalties caused by different cultural backgrounds. It was anticipated that a test similar to this culture-fair test could be used to supplement the present ASVAB causing the new total test battery to be less discriminatory against culturally different backgrounds.

II. PREVIOUS CULTURE-FREE TEST ATTEMPTS

There have been numerous attempts at devising culture-free devices for testing individuals and predicting various qualities or outcomes. Although the approaches have been quite different, the goal has been similar: to arrive at some measure of a person's abilities in such a way as to not allow that person's cultural situation to be a deciding factor in the outcome.

The Porteus Maze Test and the Navy Maze Test (Cory, 1971) attempted to measure learning ability, especially in mental group IV personnel. The purpose was to help select a portion of those who had failed standard service tests to be admitted into the military. The testee had to trace his way out of or into different mazes and was graded on various procedures and responses as well as successful completion.

Glickman, et al. (1971) examined many tests attempting culture-fair prediction of performance in such widely varying groups as taxi drivers, college students, insurance salesmen, and sewing machine workers. Training simulators, psychomotor tests, opinion/personality/interest questionnaires, and work sampling were investigated. Unfortunately, it was found that even these less traditional devices had either low predictive validity or were biased toward racial groups.

Siegel and Leahy (1974) describe an attempt to predict performance in the fleet by a type of job sampling. They

were interested in measuring learning ability on small tasks related to the machinists mate rating. Their work produced quite encouraging results, however the technique would prove hard to use with "mass production" selection processes. The time required to screen one individual would be prohibitive. Of significance is the fact that they proved that a test of learning ability can be used effectively as a nonbiased predictor of job performance.

Cory, et al. (1973) reported an interesting attempt at testing recruits undergoing basic training. The recruits were tested on their ability to perform tasks they had recently learned. The test scores were then correlated with GCT, ARI, and AFQT scores. The experiment was discontinued because it was determined that it was not possible to get enough of a distribution in scores on the task performance test. What was of interest was that there was low correlation between the Recruit Training Test and the more standard GCT/ARI and AFQT, suggesting that the two types of tests measured different abilities.

Tests which measure ability to learn such as those of Siegel and Leahy and Cory seem to hold the most promise. Unfortunately they suffer from a time and equipment intensiveness which limits efficient use on a large scale. What is needed is a measure of learning ability which could be taken in a short period of time to be compatible with present Armed Forces Examining and Entrance Station (AFEES) procedures and time constraints.

III. THE ARIMA-YOUNG TEST

Young (1975) and Arima (1978) described a nonverbal discrimination learning (DL) test[1] they constructed and administered to a group of recruits at the Naval Training Center (NTC), San Diego, California. It was decided that a nonverbal format would reduce the effect of previous educational experiences and cultural background. The specific shapes came from work done by Arnoult (1954, 1956) and are shown in Figure 1. These shapes were formed from computer generated random numbers indicating points which were then connected by lines. Shapes of this type were preferable to profiles of objects because of the culture-free goal.

Discrimination learning (DL) was chosen as the format because of its possible potential for measuring general learning ability as opposed to some more specific ability as would be measured in the more common employment testing patterns. Work sampling, questionnaires, psychomotor tests and the like could be expected to have somewhat higher validity within one specific occupation. What is needed, though, is a test which could measure a more universal quality. The armed forces have too many, quite different, job requirements

---

[1]DL is, simply, learning to discriminate a "correct" choice from a group of two or more items. It is usually done by repeated trials and errors with some reinforcement procedure used when the desired choice is made. As more and more "correct" choices are made, learning can be shown to be taking place.

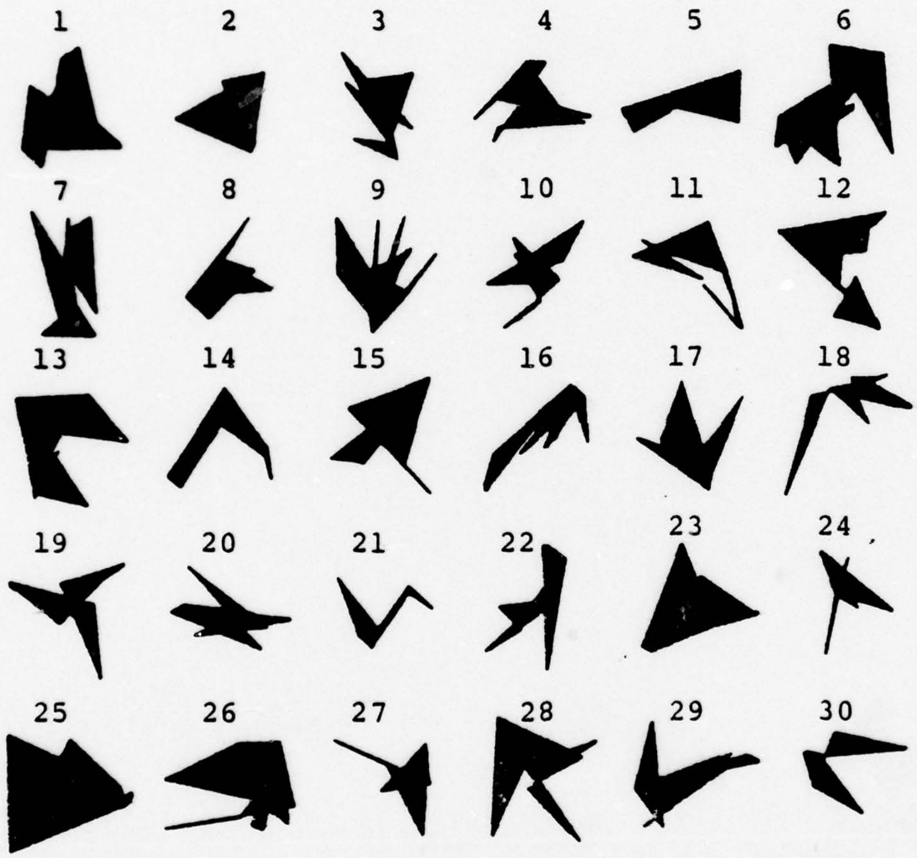


Figure 1. Shapes selected for use in assembling stimulus lists.
(From Arnoult, 1956)

to benefit from excessively specific employment aptitude tests for selection purposes.

The DL test has a final qualification which is useful in the end requirement of screening potential recruits for the armed forces. A DL type of test can be administered in a very short period of time. This characteristic becomes quite important when one considers that prospective recruits have already a fairly full schedule of mental and physical examinations during their one day at the AFEES (Navy Recruiting Manual, 1978).

The procedures used for test construction and preliminary experiments are well described by Young (1975) and Arima (1978) and will not be detailed here. The tests they devised were administered to 160 male U.S. Navy recruits at NTC, San Diego in order to evaluate the characteristics of the test under conditions as close to operational as possible and to investigate the appropriateness of the various test parameters such as scoring technique, composition, and presentation methods. Among the results attained were the following: (Young, 1975; Arima, 1978)

1. Learning did take place during the test.
2. A self-paced mode of test administration appeared to give more useful information than a machine-paced mode. It did not restrict a superior performer.
3. No significant difference was seen between white and nonwhite performance on the self-paced test.

4. Performance did not correlate highly with performance on standard verbal-oriented intelligence tests (at that time, the General Classification Test (GCT).

The research which follows was designed to begin where this previous work left off.

IV. EXPERIMENTAL GOAL

In the time since recruits from the Naval Training Center were administered the Arima-Young test there has been a major change in the test batteries used for selection. At that time potential enlistees were given the Armed Forces Qualification Test (AFQT) and some sort of supplemental test battery which varied in design and use from service to service.

Since then, all services have adopted a uniform testing program for selection and classification of most new recruits. Today each volunteer is given the Armed Services Vocational Aptitude Battery (ASVAB) which consists of twelve subtests, described in Table 1. Some special selection tests are still administered for speciality areas -- nuclear power ratings in the Navy, for example. Scores from three of the ASVAB subtests -- Work Knowledge, Arithmetic Reasoning, and Space Perception -- are combined into a general scale of mental ability. This combined score is known as the AFQT, a name held over from when there was a dedicated test by that name as mentioned above. The new AFQT score is used to separate volunteers into various mental groups, and it is the primary means of determining mental eligibility for military service. Scores from all subtests are used for classification purposes once basic mental eligibility has been established.

The purpose of the research described here was to compare the Arima-Young test with the present ASVAB in an attempt to

determine if their approach to culture-free testing measured some different capability than that measured by the ASVAB. In order for this new type of test to be useful, it must be capable of identifying some new group of individuals who have the potential to succeed in military service.

TABLE 1

Subtests of the ASVAB Form 5

Name of Test	Number of Items	Testing Time in Minutes
(GI) General Information	15	07
(NO) Numerical Operations	50	03
(AD) Attention to Detail	30	05
(WK) Word Knowledge	30	10
(AR) Arithmetic Reasoning	20	20
(SP) Space Perception	20	12
(MK) Mathematical Knowledge	20	20
(EI) Electronic Information	30	15
(MC) Mechanical Comprehension	20	15
(GS) General Science	20	10
(SI) Shop Information	20	08
(AI) Automotive Information	20	10

V. TEST DESCRIPTION

A. TEST CONSTRUCTION

Six pairs of figures were used, selected from those shown in Figure 1. They were the same as Arima and Young's Stimulus List 1 which was used in the self-paced phase of their experiment. This list contained figures previously determined to have the least similarity between figures in a pair and also the least similarity between pairs (Young, 1975). Figure 2 depicts the figures used.

One figure from each pair was choosen to be the correct response. The role of a die was used to make this assignment which resulted in different correct figures than on the Arima-Young test. Order of presentation of the pairs was the same for both tests.

The test was designed so that one pair of figures at a time was revealed. The subject chose which figures he or she felt was the right response (guessing at first) by pressing a clear panel covering that figure. If the correct choice was made, the test apparatus would cycle to the next pair. If the choice was incorrect, nothing would happen until the subject chose the other alternative. The above correctional procedure was the only reinforcement used. The six pairs were each shown ten times in random order for a total of 60 frames. Table 2 shows the order of pair presentation. The entire test was self-paced.

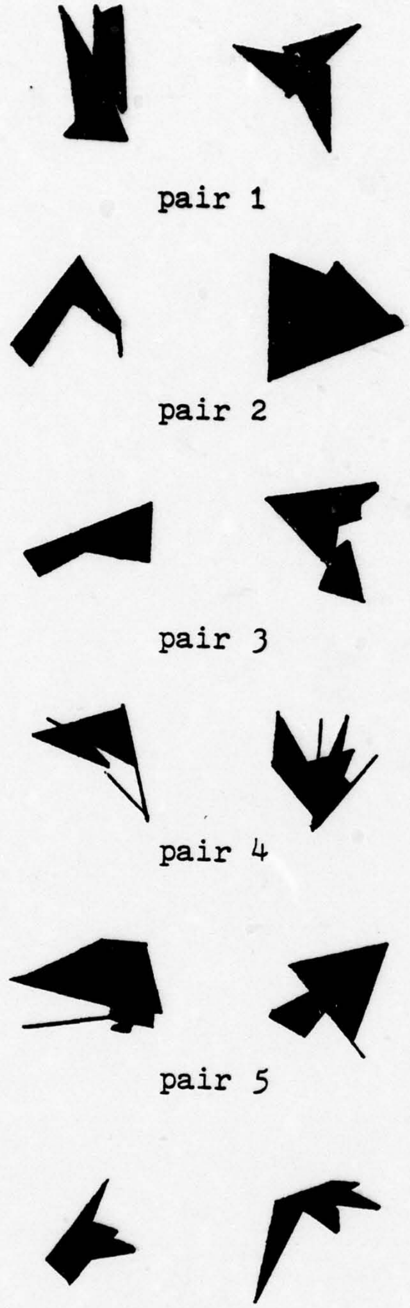


Figure 2. Test Figures

TABLE 2

Order of Stimulus Set Presentation

Repetition	Figure Order
1	2; 1; 5; 6; 3; 4
2	1; 2; 6; 4; 3; 5
3	6; 2; 5; 3; 1; 4
4	2; 3; 6; 4; 1; 5
5	4; 5; 2; 3; 6; 1
6	5; 2; 6; 4; 3; 1
7	4; 5; 6; 3; 2; 1
8	3; 6; 1; 2; 5; 4
9	6; 1; 2; 4; 5; 3
10	2; 4; 5; 3; 1; 6

Note. Figure numbers refer to those shown in Figure 1.

Other constraints used in the construction of the test were:

1. The same two figures were always displayed together.
2. The left-right order was changed randomly with the constraint that each figure appeared on the left and right an equal number of times.
3. Order within each series of six was varied.
4. All six pairs were presented before one was repeated.
5. The same pair was never presented back-to-back.
6. All figures retained the same "upright" orientation -- i.e., they were never turned around.

B. TEST HARDWARE

The equipment used, shown in Figure 3, was capable of presenting one pair of figures at a time to the subject, determining correct and incorrect responses, providing total figure exposure time, and counting total correct and incorrect responses. All equipment was manufactured by Behavior Control Institute (BCI).

The system centered around the BCI Stimulus Response Programmer. The programmer was set to operate in the multiple choice mode which was capable of displaying four choices under four clear panels (channels A, B, C, and D) on the top of the machine. A choice was made by depressing one of these panels which caused various electrical circuits to be completed depending on how the machine was programmed. Since only two choices were possible in the present test, the panels corresponding to channels A and D were deactivated and blocked.

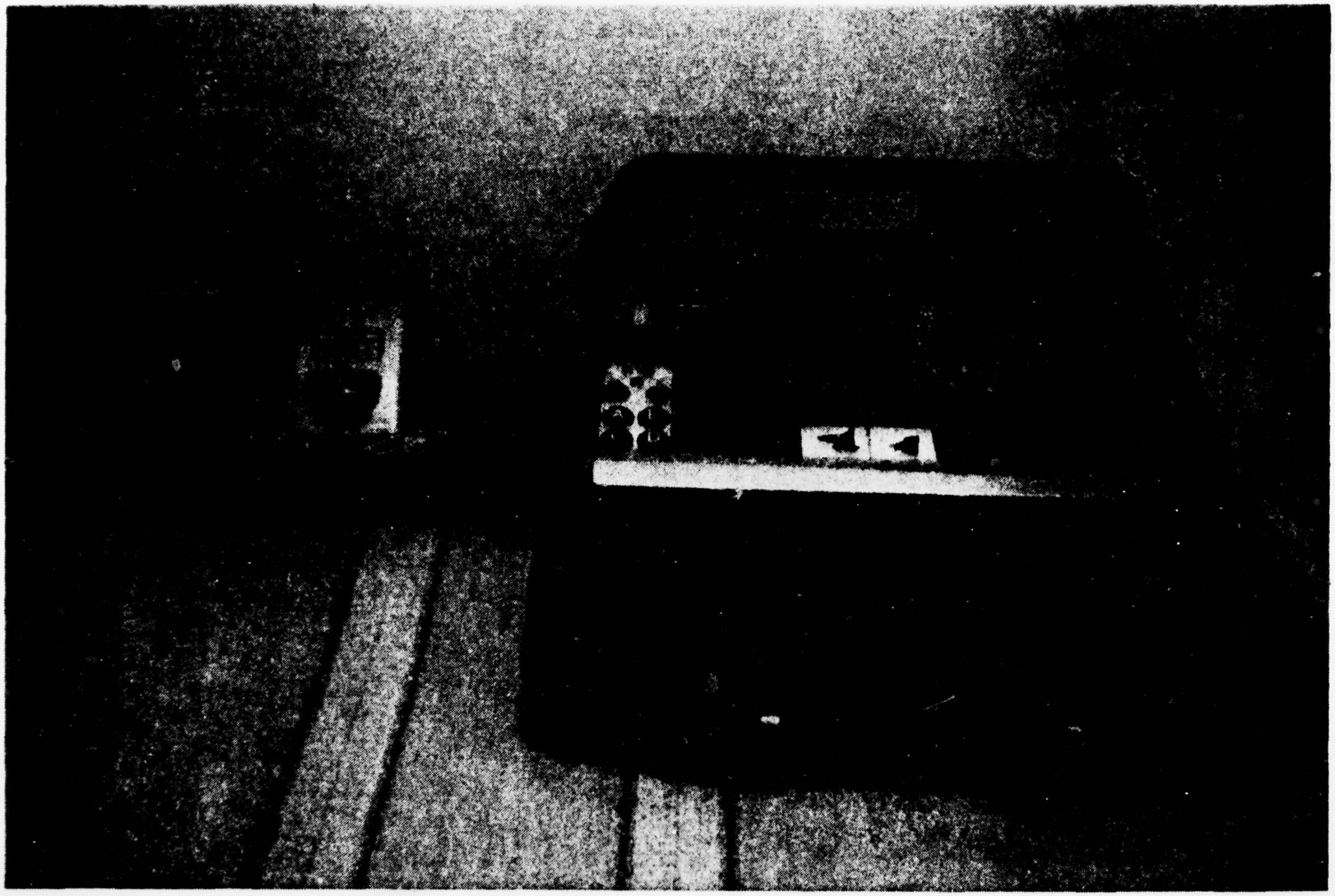


Figure 3
Test Apparatus

Machine programming was accomplished with the use of the same block of fan-folded paper that displayed each pair of figures. As the paper cycled through the machine exposing one pair at a time and stopping until correct selection was made, the machine "read" a code punched in one side of the paper which indicated the correct response and initiated movement. The machine stopped automatically at the end of the test.

A BCI Four Choice Auxiliary Control Console acted as an interface between the programmer and a BCI Scoring Indicator which cumulatively tallied correct and incorrect responses. A timer was added to the system and mounted on the Four Choice Auxiliary Control Console and electrically interfaced with it in such a way that the timer ran only during actual figure exposure time. It would stop when the paper was cycling to a new pair.

An electronic counter was added, built into the rear of the Scoring Indicator. It was designed to activate a buzzer when six correct responses in a row were made by a subject. It became unreliable during the experiment and was not used.

C. SCORING

Scoring was done by calculating an Information Processing Rate (IPR) for each subject. Each frame was considered to contain one bit of information: the correct figure of a displayed pair. A subject processed one bit of information

when he or she picked the correct figure. Therefore a rate of information processing could be computed by dividing total correct responses by total exposure time, both of which could be read directly from the test apparatus. IPR results discussed and displayed have been multiplied by 1000 in order to allow easier manipulation and discussion.

A second means of scoring had initially been planned. It was hypothesized that another way to measure performance was to record how many frames were presented before a subject scored six correct responses in a row, indicating he or she had "learned" all six pieces of information -- a trials-to-criterion measure. The equipment used to signal when six correct in a row was achieved malfunctioned early in the administration of the test to subjects, however, and this secondary method of scoring could not be used. It was discovered while this scoring system still worked that most subjects still missed many responses subsequent to scoring six in a row right. Thus it appears that a simple count of the number of frames required to score six in a row may have been an inaccurate measure of learning.

It is apparent that no real information processing took place for the first six frames since the subject had never seen the pairs before and could only guess which figure in each pair was correct. It was felt that no advantage was given to any particular individual because of this fact, and

in the interest of administration and scoring simplicity results of the first six frames were not and could not be separated from the results on the next 54 frames. All scores would have been slightly higher had only the last 54 frames been used to compute IPR.

VI. TEST ADMINISTRATION

A. SUBJECTS

In order to most accurately compare ASVAB with the Arima-Young test it was necessary to locate subjects who could take both tests prior to any selection process for military service. Only in that way would the lower spectrum of ability (as measured by the ASVAB) be represented for each test. For that reason it was not possible to administer the Arima-Young test at Recruit Training Centers. Only subjects who had already scored sufficiently high on the AFQT would have been represented. Even volunteers tested at the AFEESs would have been preselected to some extent and would have presented a biased sample.

As part of its recruiting effort the military administers a High School Testing Program. A form of the ASVAB is given to high school students as a counseling device for school counselors. The test is available to all students. While no direct recruiting takes place during the testing process, the test results are later made available to recruiters, and the entire program has proven to be a valuable recruiting tool (Navy Recruiting Manual, 1978). It was this group of high school students which seemed to offer the best medium for comparison of the two tests.

Three schools on the Monterey Peninsula cooperated in allowing this experiment. Pacific Grove High School, Seaside

High School, and Alisal High School in Salinas made available a total of 65 students who had taken the ASVAB Form 5 within the past four months. Table 3 describes the sample by race and sex.

The nonwhites were represented most heavily by individuals of Spanish descent (11 total). Only one black took the test. The rest of the nonwhite group of students is divided as follows: Filipino descent (2); Oriental descent (4); Native American descent (1); "Other" backgrounds (3). Racial group was determined by self report by each subject.

Students came from grades 9 through 12 with the average grade being 10.7. Ages represented were 14 through 18, averaging 16.2.

B. PROCEDURES

Subjects were initially given the test instructions in groups of four or less. Instructions were read to them concerning the type of test they would take, what was required, how they could determine a correct answer, and other characteristics of the test. The test instructions appear in Appendix A. Subjects were then shown an example of what they would encounter in the form of a sample test consisting of only two pairs. The figures used were from Arnoult's (1956) collection shown in Figure 1 and were not the same ones used in the actual test. The sample test served to acquaint the students with the mechanics of the test apparatus and to help illustrate

TABLE 3

Summary of Student Participation

SUBJECT GROUP	SCHOOL			TOTAL SUBJECTS
	Seaside	Pacific Grove	Alisal	
WHITE	(17)	(24)	(2)	(43)
Male	7	10	0	17*
Female	10	14	2	26**
NONWHITE	(5)	(10)	(7)	(22)
Male	2	7	2	11*
Female	3	3	5	11**
ALL SUBJECTS	22	34	9	65

*Total males - 28
**Total females - 37

the verbal instructions. Any final questions the group had were answered during and after the sample test.

Individuals took the actual test alone. Last minute questions were answered, and the test was started and continued until all 10 repetitions were presented. The total figure exposure time ranged from 35.5 sec. to 161.1 sec. The mean exposure time was 79.1 seconds. Including a 1.4 second cycle time between pairs, the entire test lasted 2.7 minutes on the average.

VII. RESULTS

Raw ASVAB scores were obtained for each subject from the schools involved in order to carry out the required analysis. A summary of scores (by mean and standard deviation) is provided in Table 4 for all subjects and followed by Tables 5 through 7 for each school. For each figure the number (N) in each category is followed by the subtest scores, the composite AFQT score and, finally, the score (IPR) from the Arima-Young test. A summary of individual IPR and AFQT scores by ethnic group is presented in Appendix B.

Zero-order, product-moment correlation coefficients were computed in order to measure the strength of the relationship between IPR and ASVAB scores. The results are shown in Table 8. The highest subtest correlation can be seen from Table 8 to be with the General Information (GI) subtest with a correlation coefficient of .34. A low but positive correlation statistically significant at the .05 level was also found for the AFQT. Figure 4 is a scattergram of individual IPR scores plotted against the AFQT score.

Subjects were then divided into two groups depending on the AFQT score. Those who scored in the top 50% nationally were separated from those who scored in the lower 50%. Again Pearson product-moment correlations were computed to determine if one group or the other had closer relations between the IPR score and the ASVAB and AFQT scores. The results are shown in

TABLE 4

MEAN TEST SCORES BY RACE AND SEX - ALL SCHOOLS

	WHITE			NONWHITE			TOTAL		
	male	female	total	male	female	total	male	female	total
N	17	26	43	11	11	22	28	37	65
GI	10.29	7.85	8.81	10.09	6.46	8.27	10.21	7.43	8.63
	1.69	1.83	2.13	2.95	2.12	3.12	2.22	1.99	2.50
WK	21.88	17.89	19.46	17.36	13.46	15.41	20.11	16.57	18.09
	5.08	5.60	5.69	7.49	6.79	7.26	6.41	6.23	6.50
MK	14.47	13.15	13.67	10.64	11.73	11.18	12.96	12.73	12.83
	4.19	4.32	4.27	4.43	5.41	4.86	4.62	4.64	4.59
GS	12.06	8.92	10.16	8.91	6.73	7.82	10.82	8.27	9.37
	4.01	3.03	3.74	2.81	3.04	3.07	3.86	3.16	3.68
NO	36.53	36.50	36.51	31.91	36.09	34.00	34.71	36.38	35.66
	7.98	8.05	7.92	9.75	11.40	10.57	8.84	9.00	8.90
AR	13.47	11.96	12.56	10.64	10.00	10.32	12.36	11.38	11.80
	4.24	3.18	3.67	4.23	3.19	3.67	4.39	3.27	3.79
EI	17.24	12.31	14.26	14.64	12.82	13.73	16.21	12.46	14.08
	5.87	4.10	5.39	4.52	2.82	3.80	5.45	3.73	4.88
SI	12.88	8.92	10.49	11.64	6.91	9.27	12.39	8.32	10.08
	3.77	2.56	3.63	3.78	2.17	3.86	3.76	2.59	3.72
AD	13.71	14.73	14.33	14.64	13.09	13.86	14.07	14.24	14.17
	3.87	3.09	3.41	3.33	4.78	4.10	3.63	3.69	3.63
SP	12.41	10.58	11.30	8.46	9.00	8.73	10.86	10.11	10.43
	5.41	3.69	4.48	3.39	4.38	3.83	5.05	3.91	4.42
MC	11.94	8.35	9.77	9.82	5.82	7.82	11.11	7.60	9.11
	3.60	3.05	3.69	2.68	1.17	2.87	3.38	2.86	3.54
AI	9.35	7.15	8.02	8.91	5.82	7.36	9.18	6.76	7.80
	4.89	3.03	3.97	2.91	2.27	3.00	4.16	2.86	3.66
AFQT	47.77	40.42	43.33	36.46	32.46	34.46	43.32	38.05	40.32
	11.36	9.73	10.89	12.45	12.36	12.28	12.87	11.03	12.05
IPR	659.06	648.00	652.37	586.64	450.00	518.32	630.60	589.14	607.00
	248.80	265.64	256.15	176.89	152.52	175.69	222.64	252.75	239.32

Note. Top number is test mean. Bottom number is test standard deviation.

TABLE 5

MEAN TEST SCORES BY RACE AND SEX - SEASIDE HIGH SCHOOL

	WHITE			NONWHITE			TOTAL		
	male	female	total	male	female	total	male	female	total
N	7	10	17	2	3	5	9	13	22
GI	9.29 1.38	7.6 1.78	8.29 1.80	9.0 4.24	5.33 2.52	6.80 3.42	9.22 1.92	7.08 2.10	7.96 2.26
WK	23.86 3.89	16.60 5.99	19.59 6.28	20.00 5.66	11.33 8.09	14.80 7.95	23.00 4.27	15.39 6.56	18.50 6.80
MK	14.71 4.11	11.50 2.76	12.82 3.64	11.00 4.24	9.33 6.11	10.00 4.90	13.89 4.20	11.00 3.58	12.18 4.02
GS	13.00 3.37	8.60 4.20	10.41 4.37	8.00 4.24	7.33 3.22	7.60 3.13	11.89 3.95	8.31 3.90	9.77 4.23
NO	36.43 9.29	36.90 8.65	36.71 8.63	36.50 6.36	37.00 14.93	36.80 11.03	36.44 8.35	36.92 9.66	36.73 8.94
AR	13.43 3.87	11.10 2.69	12.06 3.33	9.50 3.54	9.67 3.51	9.60 3.05	12.56 3.97	10.77 2.80	11.50 3.36
EI	19.00 8.08	10.90 5.22	14.24 7.53	13.50 0.71	14.00 2.65	13.80 1.92	17.78 7.41	11.62 4.84	14.14 6.63
SI	13.71 4.86	7.70 2.31	10.18 4.60	11.00 1.41	6.00 2.65	8.00 3.39	13.11 4.40	7.31 2.39	9.68 4.38
AD	13.43 4.86	13.40 3.47	13.41 3.95	14.00 5.66	12.33 4.73	13.00 4.47	13.56 4.67	13.15 3.60	13.32 3.97
SP	13.57 5.91	10.50 3.31	11.77 4.66	9.50 0.71	9.00 6.08	9.20 4.32	12.67 5.43	10.15 3.85	11.18 4.62
MC	11.86 4.56	6.90 2.08	8.94 4.07	7.50 0.71	6.00 1.00	6.60 1.14	10.89 4.40	6.69 1.89	8.41 3.73
AI	9.00 4.08	5.70 2.58	7.06 3.58	10.00 1.41	5.33 1.53	7.20 2.86	9.22 3.60	5.62 2.33	7.09 3.37
AFQT	50.86 11.12	38.20 7.77	43.41 11.03	39.00 9.90	30.00 16.52	33.60 13.61	48.22 11.51	36.31 10.19	41.18 12.07
IPR	479.00 195.57	660.00 251.06	585.47 241.31	428.50 109.60	348.67 104.29	380.60 101.75	467.78 175.17	588.15 260.24	538.91 232.51

Note. Top number is test mean. Bottom number is test standard deviation.

TABLE 6

MEAN TEST SCORES BY RACE AND SEX -
PACIFIC GROVE HIGH SCHOOL

	WHITE			NONWHITE			TOTAL		
	male	female	total	male	female	total	male	female	total
N	10	14	24	7	3	10	17	17	34
GI	11.00	8.14	9.33	10.43	7.33	9.50	10.77	8.00	9.38
	1.56	1.96	2.28	3.10	2.52	3.17	2.25	2.00	2.52
WK	20.50	19.00	19.63	18.71	17.33	18.30	19.77	18.71	19.24
	5.54	5.67	5.55	8.02	9.71	8.02	6.50	6.19	6.27
MK	14.30	14.86	14.63	10.57	11.00	10.70	12.77	14.18	13.47
	4.45	4.64	4.47	5.26	6.93	5.40	5.00	5.08	5.02
GS	11.40	9.00	10.00	9.43	8.33	9.10	10.59	8.88	9.74
	4.45	2.25	3.48	3.05	4.73	3.38	3.95	2.64	3.42
NO	36.60	35.50	35.96	28.57	27.00	28.10	33.29	34.00	33.65
	7.46	7.52	7.35	9.93	11.79	9.86	9.21	8.63	8.80
AR	13.50	12.71	13.04	10.00	10.67	10.20	12.06	12.35	12.21
	4.70	3.43	3.93	4.80	4.04	4.37	4.92	3.50	4.21
EI	16.00	13.71	14.67	15.00	14.00	14.70	15.59	13.77	14.68
	3.68	2.76	3.31	4.40	1.73	3.71	3.89	2.56	3.37
SI	12.30	10.21	11.08	12.00	7.67	10.70	12.18	9.77	10.97
	2.95	2.16	2.67	4.76	3.22	4.67	3.66	2.46	3.31
AD	13.90	15.43	14.79	14.86	14.00	14.60	14.29	15.18	14.74
	3.28	2.41	2.84	2.97	6.25	3.84	3.10	3.15	3.11
SP	11.60	11.00	11.25	8.86	9.00	8.90	10.47	10.65	10.56
	5.19	4.17	4.52	3.67	4.58	3.70	4.71	4.17	4.38
MC	12.00	9.36	10.46	10.57	6.00	9.20	11.41	8.77	10.09
	3.02	3.34	3.41	2.94	1.00	3.29	2.98	3.31	3.38
AI	9.60	8.64	9.04	8.71	7.33	8.30	9.24	8.41	8.82
	5.58	2.74	4.08	3.04	2.89	2.91	4.60	2.72	3.75
AFQT	45.60	42.71	43.92	37.57	37.00	37.40	42.29	41.71	42.00
	11.59	11.13	11.16	14.71	17.52	14.58	13.16	12.00	12.41
IPR	785.10	651.21	707.00	643.14	457.67	587.50	726.65	617.06	671.85
	204.15	295.96	265.26	160.94	99.71	165.85	195.82	279.63	244.12

Note. Top number is test mean. Bottom number is test standard deviation.

TABLE 7

MEAN TEST SCORES BY RACE AND SEX - ALISAL HIGH SCHOOL

	WHITE			NONWHITE			TOTAL		
	male	female	total	male	female	total	male	female	total
N	0	2	2	2	5	7	2	7	9
GI		7.00 1.41	7.00 1.41	10.00 2.83	6.60 1.82	7.51 2.51	10.00 2.83	6.71 1.60	7.44 2.24
WK		16.50 2.12	16.50 2.12	10.00 2.83	12.40 4.39	11.71 3.95	10.00 2.83	13.57 4.20	12.78 4.09
MK		9.50 4.95	9.50 4.95	10.50 3.54	13.60 4.61	12.71 4.31	10.50 3.54	12.43 4.72	12.00 4.36
GS		10.00 1.41	10.00 1.41	8.00 0	5.40 1.52	6.14 1.77	8.00 0	6.71 2.63	7.00 2.35
NO		41.50 12.02	41.50 12.02	39.00 9.90	41.00 7.14	40.43 7.16	39.00 9.90	41.14 7.62	40.67 7.53
AR		11.00 4.24	11.00 4.24	14.00 0	9.80 3.27	11.00 3.37	14.00 0	10.14 3.24	11.00 3.28
EI		9.50 3.54	9.50 3.54	14.50 9.19	11.40 3.21	12.29 4.82	14.50 9.19	10.86 3.13	11.67 4.53
SI		6.00 0	6.00 0	11.00 1.41	7.00 1.41	8.14 2.34	11.00 1.41	6.71 1.25	7.67 2.24
AD		16.50 4.95	16.50 4.95	14.50 4.95	13.00 5.05	13.43 4.65	14.50 4.95	14.00 4.90	14.11 4.60
SP		8.00 0	8.00 0	6.00 4.24	9.00 4.36	8.14 4.22	6.00 4.24	8.71 3.59	8.11 3.66
MC		8.50 3.54	8.50 3.54	9.50 2.12	5.60 1.52	6.71 2.43	9.50 2.12	6.43 2.37	7.11 2.57
AI		4.00 0	4.00 0	8.50 4.95	5.20 2.28	6.14 3.19	8.50 4.95	4.86 1.95	5.67 2.92
AFQT		35.50 6.36	35.50 6.36	30.00 7.07	31.20 8.35	30.86 7.43	30.00 7.07	32.43 7.59	31.89 7.11
IPR		565.50 207.18	565.50 207.18	547.00 264.46	506.20 189.86	517.86 189.97	547.00 264.46	523.14 178.96	528.44 181.31

Note. Top number is test mean. Bottom number is test standard deviation.

TABLE 8

CORRELATIONS OF IPR WITH ASVAB SUBTESTS

	GI	WK	MK	GS	NO	AR	EI	SI	AD	SP	MC	AI	AFQT
IPR	.34	.25	.20	.21	.00	.18	.03	.23	.08	.19	.28	.22	.26
	.01	.04	.11	.10	.97	.15	.82	.06	.50	.13	.02	.08	.04

Notes. 1. Top number is correlation coefficient. Bottom number is level of significance.

2. N = 65

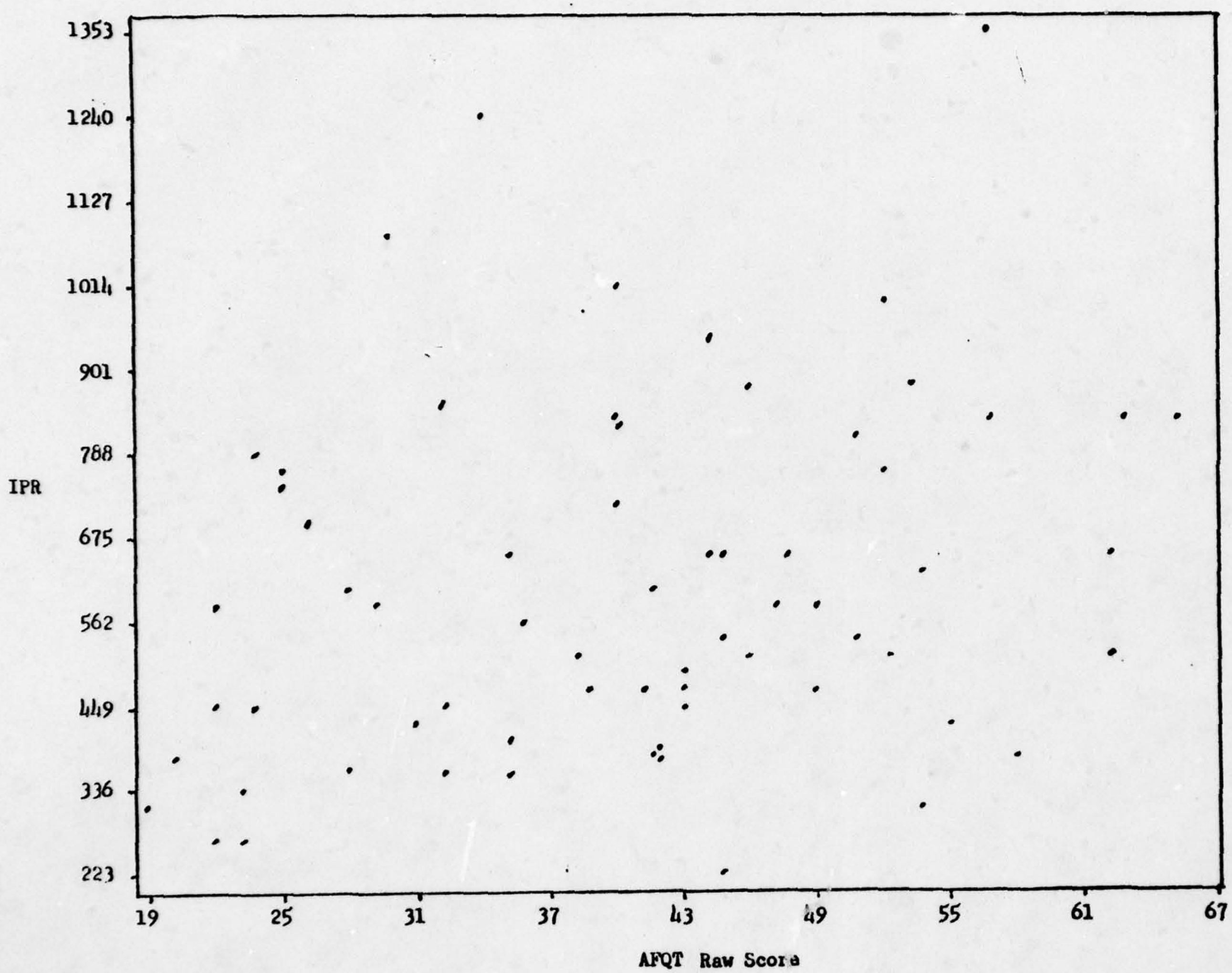


Figure 4. Scattergram of IPR Score Plotted Against AFQT Raw Score

Table 9. Both groups had similar nonsignificant correlations for the AFQT, and the only significant correlations occurred on the General Information (GI) and General Science (GS) subtests for the lower 50% group. The relative absence of significant correlations is due, in part, to the smaller sample sizes.

The sample was also divided to determine how the IPR of whites and nonwhites correlated separately with the ASVAB and AFQT scores as shown in Table 10. Again neither correlation with AFQT was significantly different from zero, and significant correlations appeared only for nonwhites on the GI, GS, and MC subtests. The difference between groups becomes more significant in this instance because the sample size is smaller for the minority group.

When males and females were considered separately (Table 11), none of the male correlations reached statistical significance at the .05 level, but the female IPR scores correlated significantly with three ASVAB subtests and the AFQT and approached significance ($p \leq .06$) on two additional subtests.

In an attempt to investigate the difference in mean IPR and AFQT scores between these various groups, a series of significance tests of the difference of two means was carried out using Student's t test for independent samples. Means for IPR and AFQT were compared between white and nonwhite, male and female, and high AFQT 50% and low AFQT 50%. The results are shown in Table 12. Significant differences on the IPR

TABLE 9

CORRELATION OF IPR WITH ASVAB BY AFQT GROUP

	GI	WK	MK	GS	NO	AR	EI	SI	AD	SP	MC	AI	AFQT
High 50%	.15	.13	.08	-.07	.14	.26	.00	.10	.18	.11	.25	.19	.26
(N = 29)	.43	.51	.70	.72	.48	.18	.99	.59	.36	.57	.19	.33	.17
Low 50%	.40	.25	.18	.37	-.15	.02	-.11	.26	-.07	.13	.24	.16	.21
(N = 36)	.01	.13	.29	.03	.37	.93	.53	.12	.71	.44	.15	.35	.22

Note. Top number is correlation coefficient. Bottom number is level of significance.

TABLE 10

CORRELATION OF IPR WITH ASVAB SUBTEST BY ETHNIC GROUP

	GI	WK	MK	GS	NO	AR	EI	SI	AD	SP	MC	AI	AFQT
White (N = 43)	.26 .09	.14 .39	.14 .38	.06 .72	-.03 .87	.09 .55	.03 .84	.19 .23	.00 .99	.14 .38	.16 .30	.27 .08	.16 .31
Nonwhite (N = 22)	.50 .02	.33 .14	.17 .46	.44 .04	-.08 .71	.18 .42	-.05 .82	.26 .25	.26 .25	.08 .73	.48 .02	-.03 .88	.27 .22

Note. Top number is correlation coefficient. Bottom number is level of significance.

TABLE 11

CORRELATION OF IPR WITH ASVAB SUBTEST BY SEX

	GI	WK	MK	GS	NO	AR	EI	SI	AD	SP	MC	AI	AFQT
Male (N = 28)	.34 .08	.00 .98	.08 .69	.20 .30	.05 .81	.11 .57	.00 .99	.14 .49	.18 .36	.14 .47	.24 .22	.10 .60	.10 .62
Female (N = 37)	.36 .03	.40 .02	.28 .09	.19 .27	-.02 .89	.23 .17	-.01 .94	.32 .05	.03 .87	.22 .19	.31 .06	.31 .06	.37 .02

<u>Note</u>. Top number is correlation coefficient. Bottom number is level of significance.

TABLE 12

STUDENT'S t TEST FOR SIGNIFICANCE OF DIFFERENCE IN MEAN SCORES

Variable	Group	N	Mean	S.D.	Standard Error	T Value	Sig. Level
AFQT	High 50%	29	51.1	6.6	1.2		
	Low 50%	36	31.7	7.5	1.2	11.30	.00
IPR	High 50%	29	648.9	236.8	44.0		
	Low 50%	36	573.2	239.3	39.9	1.27	.21
AFQT	Nonwhite	22	34.5	12.2	2.6		
	White	43	43.3	10.9	1.7	-3.00	.00
IPR	Nonwhite	22	518.3	175.7	37.5		
	White	43	652.4	256.2	39.1	2.20	.03
AFQT	Male	28	43.3	12.9	2.5		
	Female	37	38.1	11.0	1.8	1.77	.08
IPR	Male	28	630.6	222.6	42.1		
	Female	37	589.1	252.7	41.6	-0.69	.49

were found only for the ethnic comparison, but the AFQT was significantly different for the AFQT grouping and the ethnic comparison. The former is trivial, of course, since the groups were divided on AFQT.

Analysis was continued to examine the relationship of the paper-and-pencil test variables with the Arima-Young test while controlling for sex, ethnic affiliation and school. Multiple regression with forward stepwise inclusion was used to determine which variables were significant predictors of the dependent variable IPR score. The stepwise inclusion process was stopped when the adjusted r^2 was a maximum. Dummy variables were assigned to represent ethnic (white - nonwhite), sex, and school variables. Since GI had been most highly correlated in earlier analysis, interaction variables were created using GI and the demographic variables to see if GI predicted equally for all groups. AFQT was also used in creating interactive variables because of its variance between groups and because of its importance in the service selection process. Table 13 lists the variables used in the regression analysis. Results of the ensuing regression appear in Table 14.

It can be seen that only two variables are of significant use in predicting IPR scores. GI serves as a predictor for all subjects. AFQT is an aid in predicting IPR only for whites.

There is a possibility that scoring the learning test might be improved by weighting accuracy more heavily. The computed IPR was a simple division of correct answers by

TABLE 13

VARIABLES USED IN MULTIPLE REGRESSION

GI	Score on GI Subtest
EI	Score on EI Subtest
WK	Score on WK Subtest
MK	Score on MK Subtest
GS	Score on GS Subtest
NO	Score on NO Subtest
AR	Score on AR Subtest
SI	Score on SI Subtest
AD	Score of AD Subtest
SP	Score on SP Subtest
MC	Score on MC Subtest
AI	Score on AI Subtest
AFQT	Composite Score
D1	1 if Seaside High School; 0 otherwise
D2	1 if Alisal High School; 0 otherwise
D3	1 if white; 0 otherwise
D4	1 if Male; 0 if female
GID1	GI x D1
GID2	GI x D2
GID3	GI x D3
GID4	GI x D4
AFD1	AFQT x D1
AFD2	AFQT x D2
AFD3	AFQT x D3
AFD4	AFQT x D4

TABLE 14

REGRESSION OF IPR ON MULTIPLE VARIABLES

Multiple r = .489 r^2 = .239

Adjusted r^2 = .189
Standard error = 215.550

Analysis of Variance	df	SS	MS	F	Sig
Regression	4	877744.55	219436.14	4.7	.005
Residual	60	2787705.45	46461.76		

Variables in Equation	B	Beta	Std Error B	F	Sig
GI	32.22	.34	11.83	7.42	.01
AFD3	3.48	.33	1.28	7.33	.01
GID1	-12.49	-.21	6.91	3.27	N.S.
EI	-8.67	-.18	6.16	1.98	N.S.
constant 384.96					

Note. Wilkinson (1979) discusses some of the weaknesses in determining the significance of r^2 in stepwise regression. In his tables, determined by Monte Carlo Simulation of the distribution of r^2, the above results would not be significant at the .05 level unless r^2 was above .31.

total time (multiplied by 1000 for manipulation). Although it was not obvious in the experiment, a subject could score a high IPR simply by going very quickly and guessing extensively. A simple means of alleviating this potential problem is to introduce a penalty into the scoring. Subtracting the number missed from the number correct prior to dividing by total time would achieve the desired result. This formula was computed for the subjects involved in the experiment and found to be fairly consistent with IPR scores computed the original way. The two scoring methods' Pearson product moment correlation coefficient equalled .89, which exceeds the .01 level of significance. From Table 15 it can be seen that accuracy becomes more important in the second scoring method and time ceases to have a significant effect on score. It can also be seen that the people who made the highest number of correct responses tended to be the ones who also worked the fastest.

A further technique in scoring may have been a still better discriminator of ability in this test. As was attempted in the early stages of the experiment, a criterion of a certain number of correct responses in a row could be used to determine test completion. A score could then be derived from either time-to-completion or total frames exposed or some combination of the two. A major advantage of this scoring method would be that the quick learner would not be penalized by having to process ten (or however many) trials. There would have to

TABLE 15

CORRELATIONS OF TEST SCORES WITH TIME AND NUMBER CORRECT

	ADJSCR	No. Correct	Time
IPR	.89	.37**	-.26*
ADJSCR		.58**	-.19
No. Correct			-.73**

Notes. 1.) IPR = Information Processing Rate = $\frac{\text{No. Correct}}{\text{Time}}$

2.) ADJSCR = Adjusted Score = $\frac{\text{No. Correct} - \text{No. Incorrect}}{\text{Time}}$

3.) *Significant at .05 level.
**Significant at .01 level.

be some upper limit of time or frames to preclude test marathons, however.

Internal reliability of the test was measured using a split-halves technique described by Bruning and Klintz (1968). Information had been taken on 35 of the subjects as to the order of correct and incorrect responses. Odd and even frames were split and Pearson's product-moment correlation was calculated between the two halves. After correction to reflect actual test length, a reliability coefficient of .77 was obtained. This compares well with a reliability coefficient of .84 obtained by Young using a similar method.

VIII. DISCUSSION OF RESULTS

It is apparent that the Arima-Young test does indeed measure some ability other than that measured by the more traditional ASVAB. The correlation of .26 between the IPR and AFQT scores shows that there is very little similarity between the two.

An individual's AFQT score, which at present determines mental qualification for military service, is more an indication of knowledge, facts, and processes absorbed by him or her up to the time of test administration. No matter what cognitive processes are involved during the test itself, learned skills are required to perform well. If an individual, for any reason, does not have these skills, he or she will not score well on the ASVAB.

The Arima-Young test apparently measures a more basic quality. Learning was shown to have taken place in the original project (Young, 1975; Arima, 1978), and an individual's ability to learn could be differentiated by an information processing rate (IPR). Their test appears to involve more specifically cognitive factors such as Memory Span, Visual Memory, and Perceptual Speed which are discussed by Harman (1975). The low correlations in Table 8 between IPR and ASVAB subtests indicate a difference in quality measured.

Since the AFQT score was only weakly associated with the IPR score, there should then be some individuals who would

not have been accepted into the military based on ASVAB results who nevertheless displayed above-average learning ability as measured by the Arima-Young test. In fact there were four subjects in the present sample whose raw AFQT scores were below the required mark of 28 (for the Navy) but who, nevertheless, received above-average IPR scores of 607. Three of these individuals were nonwhite.

It can also be seen (Figure 4) that there was another group which did quite well on the AFQT yet scored below average on the IPR. It is hypothesized that these may represent people who make up for a lack of superior ability by hard studying -- the so-called overachievers. Hard study may be effective for paper-and-pencil tests covering traditional subjects; the Arima-Young test does not reward such behavior.

The point is not that the Arima-Young test is better than the ASVAB for service selection purposes, only that it may measure a more specific, basic ability which is less dependent on previous schooling. There must obviously still be some sort of selection device such as the ASVAB.

The military carries on much of its instruction for various specialties in traditional classroom environments with success heavily dependent on previously developed classroom skills. It would not be practical for the services to attempt to bring all personnel up to the point of verbal and mathematical expertise which are assumed to be held by students entering

some of these schools. Therefore some form of testing is required to select those who have achieved the skills necessary to do well in continued classroom training. The ASVAB is of some use in this respect (Fischl, et al. 1977).

A test of learning ability becomes important for those jobs where less traditional forms of instruction are or could be used, those jobs where one learns by watching and doing. On-the-job training requires less of the formal, school skills while still requiring an ability to absorb new knowledge.

A test such as the Arima-Young test could identify, from those doing poorly on the ASVAB, the people who still have high potential to learn on the job if accepted into the military. Its potential is as an auxiliary selection tool in a time of restricted supply.

There are certainly many other measures which could be taken to insure an adequate enlistment rate into the armed forces. Increasing retention, lowering physical standards, admitting more women, becoming more capital intensive, or reinstituting the draft would all accomplish the same thing, all with a different group of people. It is not the goal here to identify the best alternatives.

For the present sample the Arima-Young test was not culture free in the sense that whites and nonwhites did not score equally. Regressions analysis brought out that there were some differences in predictors between the two ethnic groups and that the AFQT was a predictor of the IPR score for only

the white subjects. This result, which is consistent with the results of the original study (Arima, 1978; Young, 1975), permits the use of a single equation to predict the IPR scores of both groups and is culture-free in this respect.

It may be that it is impossible to devise a test of any kind useful for service selection purposes that does not have an adverse cultural impact. Why performance on the present test was not equal for both ethnic groups can only be conjectured. The test was still a test, and those more experienced with tests may still have an advantage. Perceptions of authority, if different for each group, may have an affect on the test results. The race of the test administrator and the test environment may be influential factors. The language of the instructions, the words and phrases used, could give an edge to one group. If there is indeed an ethnic difference in IPR scores for the Arima-Young test that transcends the present sample for high school students, it would be interesting to follow age groups down through the primary grades to determine if this difference is inherent or if it is in some way developed or learned.

In order to judge its final potential, the Arima-Young test must first be submitted to a test of its predictive validity against some final criterion. The immediate problem of course is that of finding any reliable and valid criterion of job performance. Supervisor ratings could immediately bring race back as an issue; there would be no final school

grades, and grades are not really a measure of performance on the job; advancement is based on paper-and-pencil tests, something the subject has already proven he does poorly. Perhaps one must settle as Lockman (1975) did on a simple measure of survival. A person is successful if he or she is still on the job some time in the future. If, somehow, the test proves to be a valid predictor of job success, then any ethnic difference in scores could still be the basis for adverse impact if it is used indiscriminately for selection purposes.

IX. CONCLUSIONS

The following conclusions may be drawn:

1. The Arima-Young test can in fact identify, from those who score poorly on the AFQT, a group of individuals who have high learning ability, at least within the narrow parameters of the present test.

2. Whites scored higher than nonwhites on the IPR test. A final determination of culture-fairness must await validation of test results with some measure of success on the job as a criterion.

3. Study of ethnic results for various age groups starting at an early age may give further insight as to the reasons for ethnic differences in score at the high school level.

4. A more useful score may be derived by subtracting incorrect answers from correct ones prior to computing an Information Processing Rate (IPR).

APPENDIX A

TEST INSTRUCTIONS

The test you will soon see takes less than five minutes. There is no passing or failing score because the test is being used for research purposes only.

You will be seated in front of a machine which will show you two figures at a time. The figures were drawn by computer and are not supposed to represent any physical objects. For each pair of figures I have picked one to be the right answer and one to be the wrong answer. You will be shown a number of pairs of these figures. For each pair I would like you to try to pick the correct figure by pressing the clear panel directly over that figure. If the machine moves to the next pair, you know you have picked the correct figure. If nothing happens, you should pick the other figure by pressing the clear panel over it. Pushing both panels at the same time causes an error.

You will see the same pairs of figures many times. The first time you see a pair you can only guess which of the two figures I have chosen as correct. After you have guessed, try to remember which figure was correct so that you can pick that one the next time you see it. The same two figures will always be together, and the correct figure will remain the same. At times I have flip-flopped the pairs so that the correct answer may be on either the left or the right side.

Although a timer is being used, this is not a timed test. Just work as quickly as you can without rushing. From time to time a buzzer may sound. Ignore it; it is simply built into the system and does not affect you.

Here is an example of what you will see:

APPENDIX B

INDIVIDUAL IPR AND AFQT SCORES

SUBJECT	ETHNIC	IPR	AFQT
1	W	588	49
2	W	840	65
3	W	882	46
4	W	534	45
5	W	475	41
6	W	223	45
7	W	506	46
8	W	446	22
9	W	400	42
10	N	261	22
11	W	833	40
12	N	321	19
13	W	1090	30
14	W	469	43
15	W	466	39
16	N	351	32
17	W	957	44
18	W	511	62
19	W	379	58
20	W	412	35
21	N	464	49
22	W	448	32
23	W	764	52
24	W	779	24
25	N	583	22
26	W	829	57
27	W	584	29
28	W	1020	40
29	W	318	54
30	N	431	55
31	N	568	36
32	W	644	44
33	W	398	42
34	W	652	48
35	N	442	43
36	W	492	43
37	N	706	26
38	N	822	63
39	W	988	52
40	N	844	40
41	W	365	28
42	N	456	24

APPENDIX B
(continued)

SUBJECT	ETHNIC	IPR	AFQT
43	W	649	62
44	W	859	32
45	W	814	51
46	W	649	35
47	W	515	42
48	W	327	23
49	W	878	53
50	W	627	54
51	W	1250	34
52	N	649	45
53	N	374	20
54	W	540	51
55	W	1352	57
56	W	575	47
57	N	360	35
58	N	763	25
59	N	599	28
60	W	712	40
61	N	274	23
62	N	514	38
63	N	381	42
64	W	419	31
65	N	734	25

Note. W = White
N = Nonwhite

BIBLIOGRAPHY

Arima, J.K., A Culture-Free Performance Test Of Learning Aptitude, U.S. Naval Postgraduate School, NPS 54-72-2, Feb. 1978.

Arnoult, M.D., "Familiarity and Recognition Of Nonsense Shapes," Journal of Experimental Psychology, 1956, 51, 269-276.

Arnoult, M.D., "Shape Discrimination As A Function Of the Angular Orientation of the Stimuli," Journal of Experimental Psychology, 1954, 47, 323-328.

ASVAB-Counselor's Manual, Vol. 1, Armed Forces Vocational Testing Group, 1974.

ASVAB-5 Counselor's Manual, U.S. Military Processing Command, July, 1976.

Attneave, F., and Arnoult, M.D., "The Quantitative Study of Shape and Pattern Perception," Psychological Bulletin, 1956, 53, 452-471.

Atwater, D.C., Alf, E.F., and Abrahams, N.M., The Unobtrusive Measurement of Racial Bias Among Recruit Classification Specialists (NPRDC 75-6), San Diego: U.S. Navy Personnel Research and Development Center, 1974.

Bayroff, A.G., and Fuchs, E.F., The Armed Services Vocational Aptitude Battery (Tech. Res. Rep. 1161), Arlington: U.S. Army Behavior and Systems Research Laboratory, 1970.

Bilinski, G.R., Standlee, L.S., and Saylor, J.C., Effects of Sending Minority Personnel Classified As Nonschool Eligible to "A" School. Part I - "A" School Achievement (Spec Rep. 74-6). San Diego: U.S. Navy Personnel Research and Development Center, 1974.

Bilinski, C.R., and Standlee, L.S., Effects of Sending Minority Personnel Classified as Nonschool Eligible to "A" School. Part II - Performance in Fleet Assignments (Spec. Rep. 75-4). San Diego: U.S. Navy Personnel Research and Development Center, 1974.

Bruning, J.L., and Kintz, B.L., Computational Handbook of Statistics, Glenview, Illinois: Scott, Foresman and Company, 1968.

Carroll, J.B., and Maxwell, S.E., "Individual Differences In Cognitive Abilities," Annual Review of Psychology, 1979, 30, 603-640.

Cooper, R.V.L., "The All-Volunteer Force: Five Yeras Later," International Security, 1978, 325, 1-16.

Cory, C.H., A Comparison of the Porteus and Navy Maze Test (Tech. Bul. STB 71-11). San Diego: Naval Personnel and Training Research Laboratory, 1971.

Cory, C.H., Rimland, B., and Thomas, E.D., The Development and Statistical Evaluation of a Recruit Training Performance Test (Tech. Bul. SRM 73-4). San Diego: Naval Personnel and Training Research Laboratory, 1973.

Curtis, E.W., Prediction of Enlisted Performance I. Relationships Among Aptitude Tests, Navy School Grades, The Report of Enlisted Performance Evaluation, and Advancement Examinations (Tech. Bul. STB 71-10). San Diego: Naval Personnel and Training Research Laboratory, 1971.

Defense Manpower Commission, Defense Manpower: The Keystone of National Security, Washington: Government Printing Office, 1976.

Estes, W.K., "Learning Theory and Intelligence," American Psychologist, 1974, 29, 740-749.

Farr, J.L., and others, Ethnic Group Membership As A Moderator in the Prediction of Job Performance (Tech. Rep. 2). Pittsburgh: American Institute for Research, 1971.

Federal Register, Vol. 35, No. 149, Aug., 1970.

Fischl, M.A., and others, Validity of the Armed Services Vocational Aptitude Battery (ASVAB) for Predicting Performance in Service Technical Training Schools (Rep. TRR 77-4). San Diego: U.S. Navy Personnel Research and Development Center, 1977.

Holtzman, W.H., "The Changing World of Mental Measurement And Its Social Significance," American Psychologist, 1971, 26, 546-553.

Lau, A.W., Lacey, L., and Abrahams, N.M., A Noncognitive Test Battery As A Prediction of Class "A" School Performance (Rep. STB 70-5). San Diego: U.S. Navy Personnel Research and Development Center, 1970.

Lockman, R.F., and Ripma, M.G., Chances of Surviving The First Year of Service: A New Technique For Use In Making Recruiting Policy and Screening Applicants For The Navy (Rep. CNS 1068). Arlington: Center for Naval Analysis, 1975.

McCormick, E.J., and Tiffin, J., Industrial Psychology. Englewood Cliffs, N.J.: Prentice-Hall Inc; 1974.

Navy Recruiting Manual, Enlisted, Naval Recruiting Command, 1978.

Nie, N.H., and others, Statistical Package For The Social Sciences (2nd edition), New York: McGraw-Hill, Inc., 1975.

Siegel, A.I., and Bergman, B.A., Nonverbal and Culture Fair Performance Prediction Procedures: Background Test Development and Initial Results, Wayne, Pa: Applied Psychological Services Science Center, 1972.

Siegel, A.I., Bergman, B.A., and Lambert, J., Nonverbal and Culture Fair Performance Prediction Procedures II. Initial Validation. Wayne, Pa: Applied Psychological Services, Inc., 1973.

Siegel, A.I., and Leahy, W.R., Nonverbal and Culture Fair Performance Prediction Procedures III. Cross Validations and Final Report. Wayne, Pa.: Applied Psychological Services, Inc., 1974.

Sinaiko, H.W., Memo to Manpower R & D Planning Committee, Navywide Retention Conference, August 22-25, 1978." 4 Oct 1978.

Stephan, R.A., Memo to the Chief of Naval Operations, "Equal Employment Opportunity in the Navy," 4 Jan 1972.

Stephan, R.A., Memo to LCDR Toney, Navy Recruiting Command, "The Relationship Between Navy Classification Test Scores and Final School Grades in 104 Class "A" Schools," May 1972.

Wilkinson, L., "Tests of Significance In Stepwise Regression," Psychological Bulletin, 1979, 86, 168-173.

Wilson, W.R., and Miller, H., "A Note On The Inconclusiveness Of Accepting The Null Hypothesis," Psychological Review, 1965, 71, 238-242.

Young, P.A., A Culture-Free Performance Test of General Learning Ability, Unpublished Masters' Thesis, Naval Postgraduate Schook, Dec. 1975.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93940	2
3.	Department Chairman, Code 54 Department of Administrative Sciences Naval Postgraduate School Monterey, California 93940	1
4.	Professor J.K. Arima Code 54Aa Department of Administrative Sciences Naval Postgraduate School Monterey, California 93940	5
5.	LT Vining A. Sherman, Jr. c/o Cincotta 18 Prospect Street Winchester, Massachusetts 01890	1
6.	Defense Logistics Studies Information Exchange U.S. Army Logistics Management Center Fort Lee, Virginia 23801	1
7.	Navy Personnel Research and Development Center San Diego, California, 92152	1
8.	Commander, Naval Recruiting Command 4015 Wilson Blvd. Arlington, Virginia 22203	1
9.	Deputy Chief of Naval Operations (OP-10) (OP-13) Arlington Annex Arlington, Virginia 20370	1 ea
10.	Office of the Assistant Secretary of Defense (MRSL, MPP) Pentagon Washington, D.C. 20301	1
11.	Office of Naval Research Psychological Sciences Division 800 N. Quincy St. Washington, D.C. 22217	1